

# A comparison between Manchester and London neighbourhoods for relocation.

## 1. Introduction

Manchester is a very multicultural and diverse city in the North of England. It has many opportunities for young people and is surrounded by many different towns and villages all of which offer a range of lifestyle and personalities. Due to the close proximity of these areas and a good travel network, it is an ideal place to maintain a work life balance while living in the comfort of a suburb and commuting into the city for work.

In the UK, many people relocate to the south east in order to live and work in London for the career opportunities and financial gains. However, London is a very big area and travel into and across the city is vastly greater in terms of time and cost than travel in the Manchester area. For this reason, it is important that anyone who relocates to London chooses the right area in which to live, in order to be happy in their local area and minimise the need for travel outside of their commute.

For this reason, this project will attempt to find similarities between the neighbourhoods in the Manchester area and those within London area, to provide a guide to anyone seeking to relocate. Neighbourhoods will be clustered based on the venues within their centres and then a list of mappings between Manchester and London created.

## 2. Data

### 2.1 Data Sources

- Foursquare location data will be used to retrieve the location and venue details within each neighbourhood.
- Data for Manchester will be taken from the following Wikipedia page [https://en.wikipedia.org/wiki/List\\_of\\_places\\_in\\_Greater\\_Manchester](https://en.wikipedia.org/wiki/List_of_places_in_Greater_Manchester) which contains details of the Boroughs and Neighbourhoods.
- Data for London will be taken from the following Wikipedia page [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London) which contains details of the Boroughs and Neighbourhoods within London.

### 2.2 Data Usage

Using the neighbourhood data from Wikipedia, the details for 500 venues within 2000 metres of the centre of each neighbourhood will be retrieved from Foursquare. This data will then be aggregated for each neighbourhood and the top 10 types of venue will be used to cluster each neighbourhood.

The resulting list of neighbourhoods and clusters will then be split between Manchester and London, providing a mapping between the different cities.

Each cluster will be examined and allocated a descriptive name and brief overview of the characteristics of the neighbourhood.

## 2.3 Data Examples

London data:

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich <sup>[7]</sup>	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham <sup>[8]</sup>	LONDON	W3, W4	020	TQ205805
Addington	Croydon <sup>[8]</sup>	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon <sup>[8]</sup>	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge <sup>[9]</sup>	ILFORD	IG2	020	TQ455895
Aldgate	City <sup>[10]</sup>	LONDON	EC3	020	TQ334813
Aldwych	Westminster <sup>[10]</sup>	LONDON	WC2	020	TQ307810
Alperton	Brent <sup>[11]</sup>	WEMBLEY	HA0	020	TQ185835

Manchester data:

Metropolitan county	Metropolitan borough	Centre of administration	Other components
	Bury	Bury	Prestwich, Radcliffe, Ramsbottom, Tottington, Whitefield
	Bolton	Bolton	Blackrod, Farnworth, Horwich, Kearsley, Little Lever, South Turton, Westhoughton
	Manchester	Manchester	Blackley, Cheetham Hill, Chorlton-cum-Hardy, Didsbury, Fallowfield, Hulme, Moss Side, Newton Heath, Northenden, Ringway, Rusholme, Withington, Wythenshawe, Longsight,
	Oldham	Oldham	Chadderton, Shaw and Crompton, Failsworth, Lees, Royton, Saddleworth

## 3. Methodology

### 3.1 Selecting Data

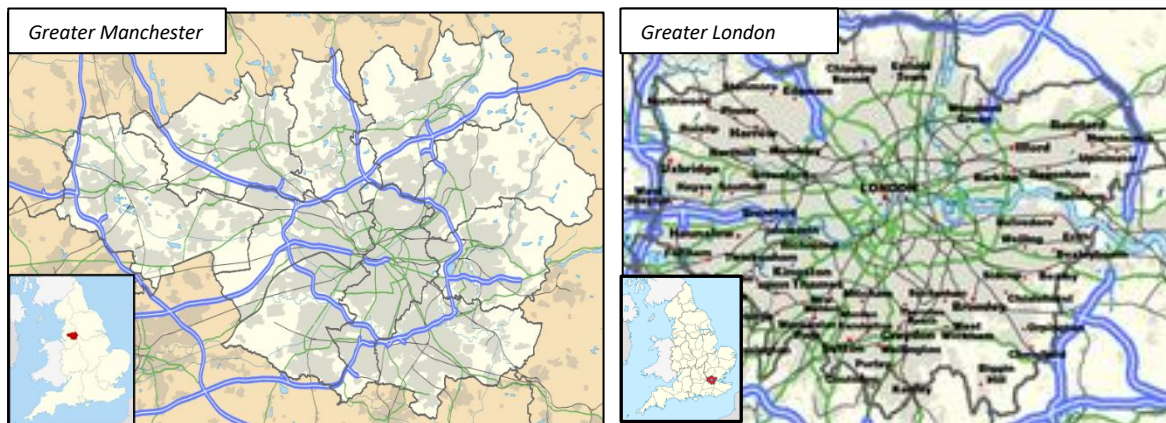
Both London and Manchester are cities that can be split in a variety of ways, including boroughs, postal areas, towns and villages. The first step was to decide what boundary I was going to use to identify my target areas and for this I relied on my own local knowledge.

Although Manchester is a defined city, I expanded my area to the Greater Manchester Metropolitan County area, which includes other towns and cities within an easy commute of Manchester.

For London, I limited my target area to the Greater London area and restricted this further to be within the M25 Motorway - which is a ringroad around the London area common. This approach is not consistent with the approach taken for Manchester, but in this case is results in a target area that is accessible by train and tube – the two most common methods of commute in the London area.

In both cases, I could have delved further into the area selection and perhaps limited it to towns and villages within a N minute commute of the cities, but I decided that this would have taken a disproportionate amount of time for a first attempt at this project.

Once I had identified my target areas I searched Wikipedia for appropriate lists of the neighbourhoods within each area. Please note that the term “neighbourhood” is very vague with regards to this project as a neighbourhood could be any of a town, village, district, borough or ward.

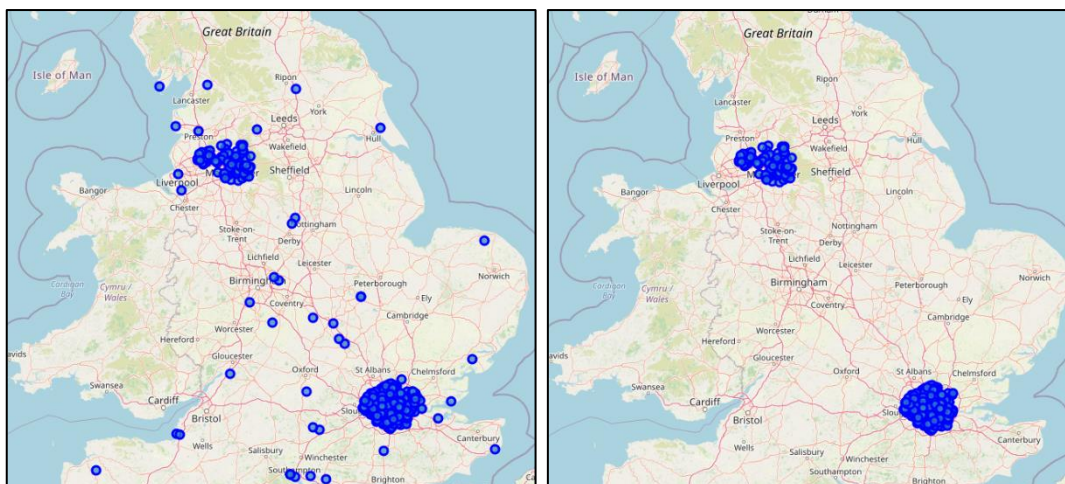


### 3.2 Cleaning the Data

Cleaning the data involved a number of steps:

1. The data was scraped from Wikipedia using `pandas.read_html()` .
2. The two datasets were then added to the same dataframe and unnecessary fields removed.
3. Geolocator was then used to identify the latitude and longitude for each neighbourhood.
4. The data was then visualised on a map using Folium.
5. Through observation it was clear that some of the neighbourhoods had not been identified correctly by geolocator. So a sample of these were checked and it was confirmed that geolocator had not been able to find the correct neighbourhood and had in fact found a similar sounding neighbourhood in a different location. There were a total of 69 neighbourhoods (11%) that were incorrect and it was deemed acceptable to remove these from the data. To do this, appropriate latitude and longitude coordinates were used to ringfence the target areas and anything outside of these zones was removed.

*Location data before and after cleaning*



### 3.3 Clustering

#### 3.3.1. Method

The method used to identify the personality and similarities between neighbourhoods was k-means clustering. This method was selected as it is unsupervised, simple and low cost. My objective for this project is to provide a guide for people looking to relocate from Manchester to London, there are many factors they must consider and to try to incorporate all possible factors would be a huge undertaking. Therefore the output of this project will be a guide to be used as a starting point of neighbourhoods they should consider relocating to.

#### 3.3.2. Parameters

Before clustering could begin, I needed to acquire the location venue data from Foursquare. To do this I used a radius of 2000m and a limit of 300 venues per neighbourhood. The reason for selecting these parameters was:

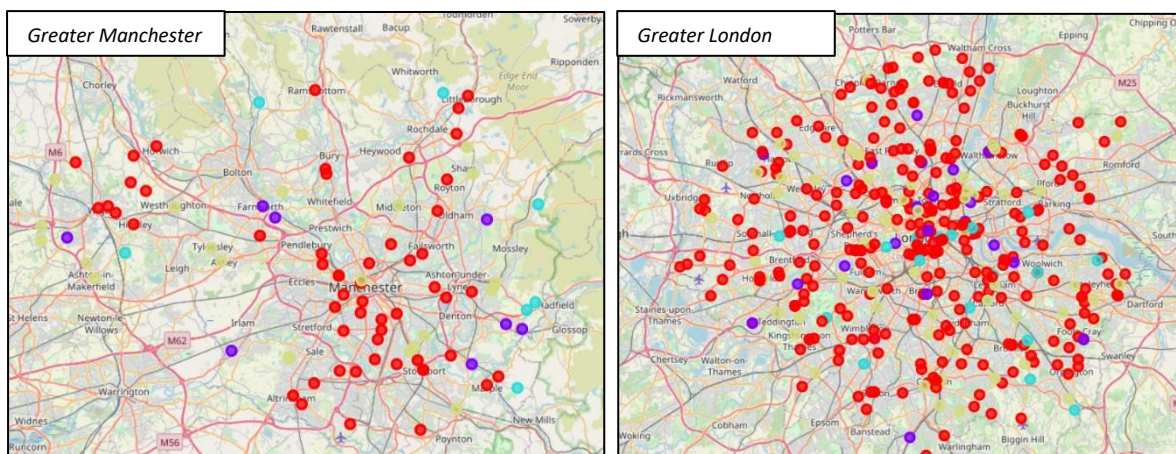
1. Radius – many of these neighbourhoods are close to each other, so rather than limit the venues to be mutually exclusive of each neighbourhood I decided that I would include all venues with walking distance of the centre. This results in a more complete picture of each neighbourhood e.g. if you live in this neighbourhood, what venues can you easily access regardless of where they are?
2. Venue limit – this was set based on the limitations of my Foursquare account. Ideally I would like to include all venues within a 2000m distance from the centre, but this would require a paid Foursquare account.
3. Final features – once the data had been acquired, it was summarised and the top 10 most common venues for each neighbourhood were selected as features.

#### 3.3.3. Clusters

The number of clusters selected was  $k=4$ . This was through a process of trial and error achieved a good spread across clusters without there being too many very small or very large clusters.

Once the clustering had been performed, they were visualised using Folium and visual inspected to make sure there was a good spread. A sample of neighbourhood “matches” were also inspected manually by comparing the top ten venues in each.

The final step was to analyse the top ten neighbourhoods in each cluster and give them a descriptive name and brief summary.





## 4. Results

The final clusters are:

### Cluster 1:

**Count:** 54 Manchester neighbourhoods, 334 London Neighbourhoods.

**Description:** Towns and villages with a good range amenities including of pubs, restaurants shops and transportation links.

### Cluster 2:

**Count:** 8 Manchester neighbourhoods, 26 London Neighbourhoods.

**Description:** Residential suburbs with public transport links, grocery shops and fast food.

### Cluster 3:

**Count:** 7 Manchester neighbourhoods, 23 London Neighbourhoods.

**Description:** Residential suburbs with local pubs and outdoor recreation.

### Cluster 4:

**Count:** 25 Manchester neighbourhoods, 82 London Neighbourhoods.

**Description:** Residential suburbs comprising of a variety pub and restaurants, with food and drink lifestyle.

The full list of mapping from Manchester neighbourhoods to London neighbourhoods can be found at the bottom of the following Jupyter notebook: <https://github.com/ninjatricky/Coursera-Capstone/blob/master/Capstone%20Final%20Notebook.ipynb>

## 5. Discussion

The results of this study have successfully provided a guide for relocation as originally intended. However, the majority of the neighbourhoods reside in only two of the clusters and it could be assumed that there is much more differentiation between these neighbourhoods than this study has uncovered. It would be interesting to rerun this study with two revisions to the methodology:

1. Include more data, such as commuting time to the city centre, population density, household income etc to give a richer picture of the neighbourhoods.
2. Use an alternative method such as hierarchical clustering to provide more distinct and varied clusters.

## 6. Conclusion

In this study I analysed neighbourhoods in the Greater Manchester and Greater London areas, to identify similar neighbourhoods based on the venues available within walking distance of each neighbourhood centre. K-means clustering was the machine learning technique used to identify similar neighbourhoods and the resulting clusters were given narrative names and descriptions. The resulting mapping of these clusters can be used by people relocating from Greater Manchester to Greater London in their search for a desirable neighbourhood in which to live.